J O I N T  C E N T E R
AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES

# The Economics of Interchange Fees and Their Regulation: An Overview

**David S. Evans and Richard Schmalensee***

**Related Publication 05-12**
**May 2005**

* Evans is Vice Chairman, LECG Europe and Visiting Professor, Faculty of Laws, University College of London. Schmalensee is Professor of Economics and Management at the Massachusetts Institute of Technology (MIT) and the John C Head III Dean of the MIT Sloan School of Management. The authors are indebted to Nese Nasif for superb research assistance, to Howard Chang, Michael Katz, Tim Muris, Jean-Charles Rochet, and conference participants for valuable comments on earlier versions of this paper, and to Visa U.S.A. for support of the research on which this paper is based. The authors alone are responsible for the opinions expressed in this paper or for any errors it may contain.

# The Economics of Interchange Fees and Their Regulation: An Overview

David S. Evans and Richard Schmalensee

## 1. Introduction

In 1958 the Bank of America began operating the BankAmericard credit card system, the predecessor of Visa, as a *unitary* system.[†] It performed both the issuing function (dealing with cardholders) and the acquiring function (dealing with merchants) itself. Similarly, it set the fees charged to both these customer classes—the annual fee, interest rate, late fees, and other fees charged to cardholders; and the per-transaction fee to merchants known as the merchant discount. It was therefore able to determine both the overall level of fees (which might be measured as total fees per dollar of transactions) and their structure (which might be measured by the shares of total fees paid by merchants and cardholders).

In 1966 the Bank of America began to bring other banks into the system as franchisees. Individual banks within the system were free then, as now, to determine the fees they charged merchants and cardholders. When a consumer holding a card issued by bank A made a purchase at a merchant that had bank A as its acquirer, bank A could, if it wished, have the same fee structure as Bank of America. But what if this same consumer made a purchase from a merchant acquired by bank B? Bank of America required the acquiring bank to pass the full merchant discount to the issuing bank. Acquiring banks had incentives to lie about their merchant discounts under this rule, as issuing banks were well aware. More importantly, this rule meant that acquiring banks received zero revenue for transactions for which they provided the merchant but had not issued the card being used. The rule therefore blunted the incentives for all banks to sign up merchants, to the obvious detriment of the system as a whole.

In 1970 the BankAmericard system was converted into a membership corporation, a *multi-party* system. This cooperative association established an *interchange fee* in 1971 to deal with transactions in which issuing and acquiring banks were different. This fee was paid by the acquiring bank to the issuing bank and initially set at 1.95 percent. It was not linked to any individual bank's merchant discount. The interchange fee thus became a revenue source on the issuing side of the credit card business and a cost element on the acquiring side. Acquiring banks

had to charge a merchant discount that was greater than the interchange fee to recover this cost. The interchange fee was an element of a standard contract that the multi-party system established for its members; other terms of the contract defined who bore the risk of fraud or nonpayment, as well as how disputes would be resolved.

This essay surveys the economic literature on interchange fees and the debate over whether interchange should be regulated and, if so, how.

## *What's Interesting About Interchange?*

Until 1979, few outside the Visa and MasterCard systems had any idea what an interchange fee was. In that year, the National Bancard Corp. (*NaBanco*) filed a lawsuit contending that when the Visa member banks determined the interchange fee, those banks engaged in illegal price-fixing, which, it was claimed, had damaged NaBanco.[‡] An appeals court found in favor of Visa in 1986, holding that the interchange fee had potential efficiency benefits for a two-sided system:

Another justification for evaluating the [interchange fee] under the rule of reason is because it is a potentially efficiency creating agreement among members of a joint enterprise. There are two possible sources of revenue in the VISA system: the cardholders and the merchants. As a practical matter, the card-issuing and merchant-signing members have a mutually dependent relationship. If the revenue produced by the cardholders is insufficient to cover the card-issuers' costs, the service will be cut back or eliminated. The result would be a decline in card use and a concomitant reduction in merchant-signing banks' revenues. In short, the cardholder cannot use his card unless the merchant accepts it and the merchant cannot accept the card unless the cardholder uses one. Hence, the [interchange fee] accompanies "the coordination of other productive or distributive efforts of the parties" that is "capable of increasing the integration's efficiency and no broader than required for that purpose."[§]

---

[†] This discussion in this paragraph and the next two follows Evans and Schmalensee 2005, pp. 153-156.
[‡] National Bancard Corporation v. Visa U.S.A., Inc. 1984, 1986a, and 1986b.
[§] National Bancard Corp. v. Visa U.S.A., Inc. 1986a, p. 602.

In 1983, William Baxter, a leading antitrust scholar who had worked for Visa on the case, published an important paper on the economic rationale for interchange fees in 1983.[**] But after the *NaBanco* decision, interchange fees faded from view in academic and policy circles and was a topic of interest mainly to industry insiders.

A few academic papers in the 1990s mentioned interchange fees,[††] but for the most part, this topic languished in obscurity until around the turn of this century. And then, as the dates on most of the entries in this essay's list of references indicate, interest in interchange fees increased dramatically among academics, banking regulators and competition authorities around the world.

Two developments caught the attention of policymakers. Cards had become an increasingly important part of the payment system in many countries. The share of consumer expenditures in the United States paid for with cards had increased from about 3 percent in 1986, the year of the *NaBanco* decision, to 25 percent in 2000.[‡‡] Similar increases occurred in other countries. For example, in 2000, cards accounted for 30 percent of consumer expenditures in Australia and 35 percent in the United Kingdom.[§§] While American Express played a significant role in the United States, globally most cards were associated with multi-party systems that had interchange fees. In the United Kingdom, for example, an influential report on the banking industry issued in 2000—the "Cruickshank Report"—addressed interchange fees in multi-party systems, concluding that "[t]here is a strong case for reform of the interchange fee system."[***]

For retailers, merchant discounts (which included interchange fees in the bank card systems) had become a growing portion of their costs, as more people paid with cards. Some retailers had periodically complained about merchant discounts; hotels had gone so far as to create their own card system in the United States in the mid 1950s to avoid the merchant discount of Diners Club (a unitary system). However, with increases in interchange fees and perhaps other legal and political developments, various organizations of retailers around the world sought regulatory relief from the fees. For example, EuroCommerce, a retailer association, filed a complaint with the European Commission in 1997. This led to an investigation of Visa Europe and ultimately a settlement in which Visa Europe agreed to lower the interchange fees.

---

[**] Baxter 1983. We discuss this paper below.

[††] These include Carlton and Frankel 1995a and 1995b, Evans and Schmalensee 1995, and Frankel 1998.

[‡‡] HSN Consultants 1987, 1991, 1992, 2001a, and 2001b, and United States Department of Commerce2003.

[§§] Visa International 2001. Measures of consumer expenditures may not be entirely consistent across countries but are sufficiently comparable to provide a rough sense of importance. See also Weiner and Wright 2005.

[***] Cruickshank 2000, p. 272.

Interchange fee levels have also been under active attack and/or regulated in a number of other countries, including Australia and the United Kingdom.[†††]

Why has interest in interchange fees surged among economists in recent years? Some, including us, were exposed to the topic through their involvement in litigation and/or regulatory proceedings. But the same could be said for many other legal and regulatory issues that have spawned much smaller literatures over longer periods of time. The large volume of theoretical literature on interchange fees has arisen for the simplest of reasons: understanding their determination and effect is intellectually challenging. As the discussion below indicates, this is not necessarily good news for policy-makers.

## *This Essay*

The remainder of this essay is organized as follows. The next section provides context by considering the operation of unitary payment systems, like American Express, in the context of the recent economic literature on two-sided markets, in which businesses cater to two interdependent groups of customers. The main focus is on the determination of price structure. We then discuss the basic economics of multi-party payment systems and the role of interchange in the operation of such systems under some standard, though unrealistic, simplifying assumptions. The key point of this discussion is that the interchange fee is not an ordinary price; its most direct effect is on price structure, not price level. While it is clear that an unregulated monopolist or a cartel in a one-sided market, like electricity generation, would set a price that is higher than would be socially optimal, no such presumption exists for the interchange fee set by a monopoly card system.

We then consider implications for privately determined interchange fees of some of the relevant market imperfections that have been discussed in the economic literature. While some studies suggest that privately determined interchange fees are inefficiently high, others point to fees being inefficiently low. Moreover, there is a consensus among economists that, as a matter of theory, it is not possible to arrive, except by happenstance, at the socially optimal interchange fee through any regulatory system that considers only costs. This distinguishes the market

[†††] See Reserve Bank of Australia 2002, European Commission 2002 and 2003, and Office of Fair Trading 2003.

imperfections at issue here for multi-party systems from the more familiar area of public utility regulation, where setting price equal to marginal cost is theoretically ideal.

The penultimate section examines the implications of the results of the previous sections for policy makers. Since there is so much uncertainty about the relation between privately and socially optimal interchange fees, the outcome of a policy debate can depend critically on who bears the burden of proof under whatever set of institutions and laws the deliberation takes place. There is no apparent basis in today's economics—at a theoretical or empirical level—for concluding that it is generally possible to improve social welfare by a noticeable reduction in privately set interchange fees. Thus, if antitrust or other regulators had to show that such intervention would improve welfare, they could not do so. This, again, is quite unlike public utility regulation or many areas of antitrust including, in particular, ordinary cartels. By the same token, there is no basis in economics for concluding that the privately set interchange fee is just right. Thus, if card associations had to bear the burden of proof—for example, to obtain a comfort or clearance letter from authorities for engaging in presumptively illegal coordinated behavior—it would be difficult for them to demonstrate that they set socially optimal fees.

We take a pragmatic approach by suggesting two fact-based inquiries that we believe policymakers should undertake before intervening to affect interchange. These inquiries are premised on the view, which we believe is now widely held, that the government should intervene in markets only when there is a sound basis for believing that it can devise policies that will improve social welfare significantly. First, policymakers should establish that there is a significant market failure that needs to be addressed. To do so they would need to examine the marginal social benefits and costs of alternative payment systems as they vary among transactions; we suggest that there is highly incomplete information available on these benefits and costs, so that any inference from the current data is at best problematic. Second, policymakers should establish that it is possible to correct a serious market imperfection, assuming one exists, by whatever intervention they are considering (such as cost-based regulation of interchange fee levels) and thereby to increase social welfare significantly after taking into account other distortions that the intervention may create. We illustrate both of these points by examining the recent Australian experience.

---

Weiner and Wright (2005) provide a useful overview of government involvement.

The final section summarizes our conclusions. Many of the results of the economic literature necessarily depend on various simplifying assumptions. In this last section we highlight those that we believe are robust, in the sense that they are likely to hold generally. Our main focus throughout is on conceptual issues; the companion paper by Weiner and Wright provides a good deal of useful factual material.[‡‡‡]

## 2. Payment Systems as Two-Sided Platforms

In the last few years, economists have come to understand that payment systems have much in common with auction houses, exchanges, shopping malls, and video game consoles. All are examples of two-sided (or, more generally, multi-sided) platform businesses. Such businesses are intermediaries that add value if and only if they can appropriately coordinate the demands of two distinct groups of customers. Beauty salons may attract both men and women, for instance, but heterosexual singles bars *must* attract both men and women—and in the right proportions. Similarly, shopping malls must attract both retailers and shoppers, auction houses need both buyers and sellers to stay in business, sellers of video games need both game players and game creators—and payment systems need both consumers and merchants.

The earliest use of the term "two-sided" in this sense of which we are aware is in a 1998 paper considering a match-making intermediary that adds value by bringing individuals of two different types together.[§§§] The authors find that, under plausible conditions, one of the types will not pay for the service, and they note that this highly asymmetric pricing is descriptive of some real two-sided markets, such as real estate agents in the United States and some dating services.

The reason why skewed pricing can happen is central to the analysis of two-sided markets in general: there are typically positive indirect network externalities between the two groups.[****] For example, the more men (women) who use a particular dating service, the more attractive the service is to women (men). If the service cannot attract both in the right proportions and sufficient numbers, it will fail. If it needs to serve men for free to accomplish this, it is rational for it to do so. Giving the service away to men, even though serving them involves positive marginal costs, can be profitable if it attracts sufficient fee-paying women.

---

[‡‡‡] Weiner and Wright 2005.

[§§§] Van Raalte and Webers 1998. While he used different terminology, Baxter 1983 did the first two-sided analysis, as we discuss below.

[****] See Ackerberg and Gowrisankaran 2003 for an attempt to quantify these externalities in a payments context.

The recent general literature on two-sided platform markets began around 2002 with early versions of a seminal contribution by Jean-Charles Rochet and Jean Tirole.[††††] Early versions of an important related paper by Mark Armstrong appeared shortly thereafter.[‡‡‡‡] These papers pointed out that many businesses or markets can usefully be thought of as two-sided, and they developed some general implications of the importance of balanced participation from the involved customer groups (or, as it is more commonly put, "getting both sides on board"). Since then the literature on two-sided platform markets has grown explosively.[§§§§] This literature generally considers one or more vendors dealing directly with both (or all) involved customer groups and it is thus most directly relevant in our context to the analysis of unitary payment systems.

## *Pricing by Two-Sided Platforms*

Almost all theoretical analyses of pricing by two-sided platforms assume that they either charge only an access or membership fee (following Armstrong) or that they charge only a variable or per-transaction fee (following Rochet and Tirole). While this simplification generally facilitates understanding of basic principles, it is problematic in some cases. Newspapers, for example, typically charge an access fee to readers (a fixed cost for the newspaper regardless of what is read), while advertisers pay a variable fee based on the number of readers. In the payment card context, for instance, merchants incur small fixed access (terminal) costs of card acceptance as well as per transaction merchant discounts. On the other side of that market, consumers sometimes pay membership fees, but variable fees are typically slightly negative as a result of free float and sometimes noticeably negative as a result of transaction-based reward programs.[*****]

[††††] Rochet and Tirole 2003b.
[‡‡‡‡] Armstrong 2004.
[§§§§§§§§] Contributions that look beyond payment systems include Armstrong 2005, Armstrong and Wright 2004, Bolt and Tieman 2004a and 2004b, Caillaud and Jullien 2003, Gabszewicz and Wauthy 2004, Hagiu 2004a and 2004b, Jullien 2004, Rochet and Tirole 2004a and 2004b, and Schiff 2003. Evans 2003a and 2003b and Wright 2004b provide accessible discussions of the implications of some of this work as well as empirical support; Rochet and Tirole 2004b provide a more technical overview.
[*****] As we note below, the payment card industry has followed this pricing structure since its birth, as a two-sided platform, in 1950.

The literature on pricing by two-sided platforms distinguishes between the level of prices and the structure of prices. The profit-maximizing *price level*—the price paid by men plus the price paid by women in the dating service example, for instance—depends on the costs of serving both groups, on the sensitivities of both demands to price, and on the indirect network effects between the two customer groups. The price sensitivities depend in the usual way on the price and quality of available substitutes and other factors, including, in some models, the presence of two-sided competitors.

The profit-maximizing *price structure*—the ratio of the price paid by men to the price paid by women, for instance—also depends in general on costs, price sensitivities, and the way participation by the members of each group affects the demand of the other group.[†††††] In addition, all general models of two-sided platform markets imply that profits may be maximized by highly asymmetric pricing in which one group is served at a price close to or even below marginal cost, and most or all gross margin is earned by serving the other group.[‡‡‡‡‡]

It is important to note that many, if not most two-sided markets exhibit this sort of asymmetry in pricing and gross margin generation.[§§§§§] Shopping malls, for instance, often provide free parking to consumers, sometimes in expensive parking structures, and make all their money by charging rent to merchants. Yellow Pages and competing telephone directories of merchants in the US are given away to consumers; all the revenue is provided by merchants.[******] Similarly, Microsoft and Apple do not charge applications software developers anything for the highly valuable software services (sometime called APIs) included in their software platforms. Both these firms make almost all of their money from end-users of computer systems. (In Microsoft's case, Windows is usually licensed to computer makers that in turn license it to end-

[†††††] In the Rochet-Tirole model, which is based on particular assumptions about the structure of demand, the price structure is completely independent of the relative costs of serving the two groups (there are no fixed costs on either side in their formulation). See Rochet and Tirole 2003b, pp. 996-997.

[‡‡‡‡‡] We define gross margin here and below as revenues minus direct, side-specific variable costs, such as the manufacturing costs of video game consoles. Attributing profits to one side or another is strictly incorrect, since both sides must be on board for profits to be earned, and many costs in platform businesses, like the cost of developing the Windows operating system, are properly thought of as joint between the two sides.

[§§§§§] Bolt and Tieman 2003 offer an explanation based on the fact that some demand functions that are well-behaved in an ordinary one-sided context yield non-concave profit functions in some two-sided models, thus making corner solutions (e.g., prices of zero) optimal.

[******] In the payment card context, Vickers 2005 has argued that one might find "a distortion of competition" when one side of the market is charged "more than all the costs" of providing services to both sides. There is no theoretical basis for this test of which we are aware, and this sort of pricing is not uncommon. It occurs whenever, as in the two examples in the text, one side of a profitable two-sided business pays nothing, so the other side must be charged "more than all the costs" if the enterprise is to be profitable.

users.) On the other hand, makers of video game consoles sell them to end users at or below cost and make most or all of their gross margin from license fees paid by game developers.[††††††]

## *Unitary Payment Systems*

The general literature on multi-sided markets has immediate application to the analysis of unitary payment systems—like the BankAmericard system before franchising, the American Express system today (ignoring recent franchise-like bank deals), or the store-specific cards offered by such merchants as Neiman-Marcus. Neglecting for the moment the distinction between access and variable prices, the price structure here can be described by the ratio of fees paid by merchants, typically in the form of the merchant discount, to transaction-related fees paid by consumers, in the form of annual and other fees.[‡‡‡‡‡‡] Using this measure, the available evidence indicates that unitary systems have generally adopted asymmetric pricing structures and earned the bulk of their revenue from merchants, rather than consumers.

Before 1950, payment cards were issued by retailers for use in their stores. Then, as now, cardholders who paid their bills within a specified amount of time (usually a bit less than a month) did not pay any fee for charging and in fact benefited from the float. Those who financed their store card charges paid interest, of course. But to our knowledge, then as now, retailers did not cancel cards held by customers who chronically did not finance. Although we would not want to push this point too far, given that store cards are one-sided and are bundled with finance services, store card transaction services have what might be thought of as a slight negative variable price and a zero access price.[§§§§§§] Merchants with retail cards presumably find that this is the optimal pricing scheme.

Diners Club introduced the first two-sided payment platform—that is, a general purpose payment card that could be used by cardholders at many retailers—in 1950. After initially

[††††††] See Evans et al. 2004.

[‡‡‡‡‡‡] By "transaction-related," we mean that we are here following almost the entire literature and focusing on the payment function of credit, charge, and debit cards, putting to one side the credit function that only credit cards perform, and the associated revenue flow from consumers.

[§§§§§§] Store cards may increase purchases by relaxing short-run liquidity constraints. As we note below, this argument has been advanced as a motivation for retailer acceptance of general purpose payment cards (particularly credit cards) by Chakravorti and To 2003. See also Chakravorti and Emmons 2003 and Wright 2000. However,

offering the cards for no fee to consumers, Diners Club settled on a business model in which cardholders paid an annual fee of $3 (over $18 in 2004 dollars) and a slightly negative transaction fee in the form of float, while merchants paid a variable fee of seven percent of each transaction. During the 1950s and into the early 1960s, based on available data, Diners Club earned about 70 percent of its revenues—and probably most of its gross margin—from the merchant side of the business.[*******]

American Express entered the card business in 1958. It adopted a similar business model, though it eventually settled on slightly lower merchant fees and slightly higher cardholder fees than Diners Club. But available data indicate that for the last four plus decades American Express has earned upwards of 65 percent of its transaction-related revenues from merchant fees (in later years some American Express cards began to bundle borrowing with transaction services, which makes clean comparisons more difficult).[†††††††] Bank of America also entered in 1958, with a card that did not charge consumers a fee, although there were finance charges (and associated costs of funds and default risk) from consumers who chose to revolve. Their merchant fee was 5 percent of each transaction.[‡‡‡‡‡‡‡]

Even though transactions-related costs on the acquiring side of the business seem to be lower than those on the issuing side,[§§§§§§§] these unitary systems apparently concluded that the profit-maximizing price structure for the system—the price structure that gets and keeps both sets of customers on board and permits both issuers and acquirers to be profitable—is one that obtains the bulk of the revenue from the merchant side of the business.

These unitary systems instituted these pricing schemes in their early years, when they would appear to have had little market power. We think it is fair to conclude that the

---

again note that we would expect that retailers would take efforts to weed out transactors if that were the entire motivation.

[*******] Evans and Schmalensee 2005, pp. 54-55. Annual fee converted to real 2004 dollars using the GDP implicit price deflator from United States Department of Commerce 2004.

[†††††††] Evans and Schmalensee 2005, pp. 57-59, 150. In 2004, about 84 percent of American Express's card transaction-related revenues came from merchants, excluding finance revenues from cardholders. The proportion is about 63 percent when net finance revenues are included and using American Express's card revenues on a "managed" basis (which includes revenues for card loans that have been securitized); this same measure is 71 percent on a GAAP basis (which excludes finance charges on securitized card loans). Figures calculated using data from American Express 2005, pp. 45-49.

[‡‡‡‡‡‡‡] Wolters 2000, p. 331.

[§§§§§§§] The portion of the merchant discount retained by acquirers on MasterCard and Visa transactions in the United States averages about 0.4 percent, which goes to cover acquirers' costs. This compares to an average interchange fee of about 1.7 percent, which, in combination with additional cardholder fees, goes to cover issuers' costs. Evans and Schmalensee 2005, pp. 11, 262.

"competitive"—certainly in the sense of non-collusive and non-monopolistic—pricing structure for payment cards is (or at least was for some time) one in which merchants pay a relatively high transaction price and cardholders pay zero or possibly slightly negative transaction prices plus modest fixed fees, and in which the bulk of the profits, loosely speaking, thus flow from the merchant side. Being the competitive pricing structure does not necessarily mean that it is socially optimal, however, and it is to that issue we turn next.

## *Profit versus Welfare*

How do prices charged in two-sided markets compare with socially optimal prices? Most theoretical analysis of this question considers a single firm with some market power (i.e., it faces demand curves that slope down—though they may be highly price-elastic) selling to two customer groups in an otherwise perfectly competitive economy. Under these assumptions it is clear that the price level will be too high because of the exercise of market power—a conclusion that, of course, applies to most firms in real economies.

In fact, the monopoly pricing problem is even more serious than usual: even if marginal costs are constant and there are no fixed costs, the first-best social optimum requires setting prices that do not cover total cost, just as is the case with natural monopoly.[********] The intuition is straightforward. In an ordinary one-sided market under the usual assumptions, buyers' willingness to pay for incremental units of output provides a measure of the social value of that output. Thus the social optimum in such markets occurs at the output level at which price is equal to marginal cost, since at lower levels of output buyers are willing to pay more than marginal cost for incremental output, while at higher levels of output they are willing to pay less than marginal cost. In a two-sided market, however, increases in output on side A of the market provide positive benefits to buyers on side B that are not reflected in the side A demand curve. Thus if price equals marginal cost to customer group A, it is nonetheless socially beneficial to increase output to A because of the (externality) benefits that would thereby be conferred on members of group B.

[********] This is explicitly shown in Armstrong 2004 and Bolt and Tieman 2004b. It is implicit in Rochet and Tirole 2003b, since their analysis of socially optimal pricing imposes the binding constraint that the firm just break even.

What about price structure? In general single-firm models of two-sided markets in otherwise perfectly competitive economies, one can compare the conditions defining profit-maximizing pricing, welfare-maximizing pricing (which involves the seller losing money and, presumably, being subsidized by the government), and Ramsey pricing (which involves maximizing social welfare subject to the constraint that the firm does not lose money). Comparing these conditions in various models yields two general observations that appear to be robust to modeling assumptions. First, neither welfare-maximizing pricing nor Ramsey pricing is ever purely cost-based. In both cases, the optimal price structure also depends on price sensitivities and externalities on both sides of the market.[††††††††] Second, there is no simple, general description of the relations among the profit-maximizing and Ramey price structures. They are rarely identical, but the sign and magnitude of the difference between them depends on essentially all the demand, cost, and externality parameters in the model.

## 3. Interchange Fees in Multi-Party Systems

Multi-party systems emerged in the mid 1960s.[‡‡‡‡‡‡‡] One was the BankAmericard franchise system that we mentioned earlier. Others were cooperatives of banks that agreed to collaborate on a card brand and, in effect, pool the merchants they had signed up so that any individual with a card from a member of the cooperative could use their card at any merchant also signed up by any member of the cooperative. The predecessor of MasterCard emerged during this period as a national cooperative of banks. BankAmericard adopted a cooperative structure a few years later.

These banks initially had a purely practical problem to solve. When a cardholder serviced by bank A presented her card to a merchant serviced by bank B, the two banks had to have agreement on many issues in order to execute the transaction, even if that meant following custom or a default rule. One set of issues concerned which party bore various risks—nonpayment by the cardholder, non-delivery of goods by the merchant, bankruptcy of the merchant or, for that matter, of the merchant's bank. Another issue was how much the issuer and

[††††††††] In fact, in the Rochet-Tirole model, which assumes a particular demand structure, the Ramsey price structure, like the profit-maximizing price structure, is completely independent of group-specific marginal costs. See Rochet and Tirole 2003b, pp. 997-998.

[‡‡‡‡‡‡‡] This discussion in this section is based on Evans and Schmalensee 2005, pp. 153-156.

acquirer were compensated for executing a transaction that could not take place without participation by both of them.

As noted above, when Bank of America began franchising, its first response to this problem was a rule stipulating that whenever acquiring bank A and issuing bank B were different, A was to send the full amount of its merchant discount to B (and, of course, B was to reciprocate when one of its merchants dealt with one of A's cardholders). Presumably the idea was that if all banks adopted similar price structures, then to a first approximation typical bank A's merchant/consumer revenue mix would not depend on whether its cardholders dealt with its merchants or with those acquired by other banks. If this had worked, the system could have maintained a merchant-centered (or any other) price structure. But, as we observed above, this device failed—both because it reduced everyone's incentives to sign up merchants and because it invited deception.

In the early cooperatives, some banks entered into bilateral agreements with each other; this was possible when the systems had few members. But ultimately the cooperatives decided to develop a default set of rules, or contracts, that defined the allocation of risks and payments. In NBI—the cooperative that evolved from the BankAmericard franchise system and was the predecessor of Visa—as long as an acquirer's merchant met certain terms, such as properly authorizing transactions and checking card numbers against a list of known fraudulent accounts, it was guaranteed payment. If a transaction turned out to be fraudulent or a consumer failed to pay, the issuer was responsible. Various procedures were also set up to resolve disputes between merchants and cardholders as to the validity of particular charges. MasterCharge—MasterCard's predecessor—developed a similar contract with a similar interchange fee.

Both contracts were presumably the result of a bargain struck within the cooperative organizations between members with different stakes in acquiring and issuing (although back then there was less specialization on this dimension than there later came to be). Given the payment card industry at the time, the price structure that resulted from these interchange fees was similar to that of the unitary systems, although the price levels were lower. Since the interchange fee was a cost to the acquirer it was passed on to merchants as part of merchant discounts. The resulting merchant discounts were lower, however, than those charged by American Express at the time. The issuers then chose to issue cards with modest access fees and

slightly negative transaction fees for cardholders.[§§§§§§§§] Unlike the charge-card systems, of course, the credit cards came bundled with longer-term financing. For transactions, the resulting percentages of revenue that the multi-party systems earned from the merchant side were similar to those of American Express and Diners Club, if not higher.

## *Alternatives to Interchange*

What might the early multi-lateral card systems have done instead? Some have argued that for competition policy reasons, bilateral negotiations should have been used to set the terms of two-bank transactions rather than collective action.[*********] This argument seems farfetched for the early cooperatives. They had no collective market power by any measure, and it seems more plausible that the interchange fee was, as advertised, devised to reduce the transactions costs of entering into bilateral negotiations. Moreover, as we noted above, the *NaBanco* court found that bilateral negotiations were not a practical solution in systems with many banks, both because of transactions costs and because the honor-all-cards gave issuers substantial leverage over acquirers.[†††††††††] As long as an honor-all-cards rule is in effect, so that merchants are required to accept all cards of a given brand, an acquirer is at a significant disadvantage in negotiations with other issuing banks, since its merchant is required to accept their cards, but it has no guarantee of payment by the card issuer. A guarantee of payment is possible only when the terms of payment, including the interchange fee, if any, are specified. Because of this asymmetry, there is no reason

[§§§§§§§§] At least through most of the 1970s, annual fees for Visa and MasterCard cards were relatively uncommon. National Bancard Corporation v. Visa U.S.A., Inc., p. 27.

[*********] E.g., this was one of the remedies proposed by NaBanco. National Bancard Corporation v. Visa U.S.A., Inc. 1984, p. 1241.

[†††††††††] A recent analysis (Small and Wright 2000) concludes that, at least under certain somewhat special assumptions, even if transactions costs were low enough to make bilateral negotiations practical in large systems, banks' strategic behavior would undermine the system's viability. It is worth noting, though, that the terms of two-bank transactions in the Australian EFTPOS system (a PIN-based debit card system) are set exclusively by bilateral negotiation; see Reserve Bank of Australia 2004. There are only 11 acquirers and 150 issuers in the EFTPOS system; these numbers, while nontrivial, are orders of magnitude lower than the thousands of issuers and (at least) hundreds of acquirers in, e.g., the Visa system in the United States. See Reserve Bank of Australia 2005a, pp. 4, 40-45. Wright and Weiner 2005 report that interchange fees are also set by bilateral negotiation in Sweden. Banking concentration is quite high in Sweden, with the top four banks accounting for over 80 percent of total assets in 2003. See Swedish Bankers' Association 2004. By contrast, the top four banks in the United States accounted for only a 24 percent share in 2002. See United States Census Bureau 2004.

to believe that bilateral negotiation would generally lead to lower average interchange fees or merchant discounts than multilateral action at the association level.

Others have argued that the interchange fee should simply have been set at zero by competition authorities or some agency. Credit card paper would then exchange "at par," like checks in the United States.[‡‡‡‡‡‡‡‡] One can raise several questions about this proposal. First, a zero interchange fee would result in lower prices for merchants (since acquirers would not have the interchange fee cost) and higher prices for cardholders (since issuers would not have this source of revenue). Although this might seem "fair" in a philosophical sense, there is no basis in the economics of two-sided industries for presuming that this pricing structure is more—or less—efficient than one that, like the structures adopted by the unitary systems, imposes higher prices on merchants. That is, without further information there is nothing economically special about an interchange fee of "0"; there is no economic basis for concluding that an interchange fee of "0" is better or worse for society than any randomly chosen positive or negative percentage.[§§§§§§§§§]

Second, setting the interchange fee at zero imposes a particular price structure on the system, one in which side-specific prices are tightly linked to side-specific costs. For example, had the predecessors of MasterCard and Visa had a zero interchange fee in the early 1970s, they would have had to raise card fees by $4.88 per account (or about $18 in 2004 dollars) to compensate for the loss of interchange fee revenues.[**********] Imposing less pricing flexibility on the emerging cooperative card systems in the early 1970s would have necessarily placed them at a competitive disadvantage relative to the more established unitary systems, which could choose their price structures without constraints. This distortion in competition between unitary and multi-party systems would need to be weighed against whatever benefits policymakers believe would arise from mandating exchange at par.

One also has to consider what other changes might result from a zero interchange fee. It is possible that a multi-party system is not even viable with the pricing structure that would result

[‡‡‡‡‡‡‡‡] See, e.g., Frankel 1998 and Balto 2000 and, for a rebuttal, Ahlborn et al. 2001.

[§§§§§§§§§] Interchange payments in the Australian EFTPOS system flow from issuers to acquirers; see Reserve Bank of Australia 2004. Negative interchange fees of this sort existed in some online debit card systems in the United States until 1997 and remain the norm in U.S. ATM systems; see Hayashi et al. 2003.

[**********] Calculated from data provided by Visa U.S.A. Card fee converted to real 2004 dollars using the GDP implicit price deflator from United States Department of Commerce 2004.

from a zero interchange fee.†††††††††† In their early days, the viability of these systems was very much an open question. Even today, large numbers of issuers might move to other card systems (possibly new ones) that are organized so as to be able to replace interchange revenue without regulatory or antitrust scrutiny. (Consider, for instance, franchise arrangements with for-profit unitary systems in which merchant discounts and payments to issuers are set unilaterally by the system.) It is not clear that the pricing structure from such systems would be more favorable to merchants. Another possible effect of a zero interchange fee would be to change the other terms of the contract among member banks. For example, rules governing disputes among issuers and acquirers might be made more favorable toward issuers in order to avoid issuer defections. Or the circumstances under which payment is guaranteed to acquirers might become more limited.

## *The Baxter Analysis*

William Baxter addressed the issue of system viability in his pioneering analysis of interchange fees.‡‡‡‡‡‡‡‡‡ His was also the first paper of which we are aware that showed an understanding of the two-sided nature of payment systems—or, indeed, of any market. As we will see, although his model is special and unrealistic in some respects, it provides important insights.

Baxter assumed perfect competition among issuers, among acquirers, and everywhere else in the economy, and, like almost all the subsequent literature, he assumed away all fixed costs and access prices. Under these conditions, collective determination of the interchange fee cannot be an exercise of market power, since there is no market power anywhere in the economy. It is simply a payment from one set of perfectly competitive firms, which will have to raise their (variable) prices to cover it, to another set of perfectly competitive firms, which will lower their (variable) prices so as to compete it away completely. The interchange fee thus can only affect the price structure, not the price level. Baxter assumed that consumers and merchants would use a particular payment card if and only if the per-transaction price charged to them was less than the per-transaction benefit from using the card rather than cash or check.

---

†††††††††† We illustrate this point in a simple model immediately below.
‡‡‡‡‡‡‡‡‡ Baxter 1983.

To see the role of the interchange fee in ensuring system viability in this setup, suppose for simplicity that all consumers have the same per-transaction benefits and that so do all merchants. Let the per-transaction prices charged by the card system to consumers and merchants be $P^c$ and $P^m$, respectively; let the corresponding per-transaction benefits relative to cash or check be $B^c$ and $B^m$; and suppose the constant per-transaction marginal cost of serving a consumer is $C^c$ and of serving a merchant is $C^m$. Under these assumptions, it is efficient to use the card for all transactions if and only if

(1) $$(B^c + B^m) \geq (C^c + C^m).$$

With a zero interchange fee and perfect competition, consumers will agree to use the card and issuers will break even if

(2) $$B^c \geq P^c = C^c,$$

And merchants will agree to accept the card and acquirers will break even if

(3) $$B^m \geq P^m = C^m.$$

For a zero-interchange system to be viable, both (2) and (3) must be satisfied. It is easy to find numerical examples in which (1) is satisfied, but either (2) or (3) is not. Suppose, for instance:§§§§§§§§§§

(4) $B^c=1$, $B^m=8$, $C^c=3$, and $C^m=2$.

Here total per-transaction benefits from using the card are almost double the corresponding cost, but there is no price to consumers that satisfies (2). In this example retailers receive the most benefits if the card is used, and it would be logical for them to cover most of the system's costs. But there is no way to accomplish this with a zero interchange fee. An interchange fee of 2 solves the problem. This raises the acquirer's cost and (because of perfect competition among acquirers) the merchant discount, $P^m$, to 4, and it lowers the issuer's cost and (because of perfect competition among issuers) the consumers' fee to 1. Merchants contribute 80 percent of the system's revenue but are better off by 4 per transaction, while consumers are just indifferent to the card's existence even though the positive interchange fee has reduced their fees substantially.

By assumption, the costs incurred by both sides are necessary to execute a transaction. If this assumption is correct, it makes no sense to think of either side as providing particular services to the other; both must incur all the costs stated for either to benefit. Moreover, there is no way for regulators to look only at cost conditions and conclude that an interchange fee of 2 is

appropriate. Even in this simplest possible example, demand conditions must be considered. It could happen, of course, that some cost-based formula produced an interchange fee of 2. But this could only happen by chance.

In Baxter's model, the interchange fee is not set to maximize profits, since there are no profits earned anywhere in the payment system. He argued that its level was determined uniquely by the need to balance the supply and demand of card transactions. This seems somewhat artificial, since as long as some merchants have agreed to accept a card, the volume of transactions is determined unilaterally by card-carrying consumers' deciding whether or not to use the card for particular transactions. Nonetheless, the proposition that at some interchange fee levels the system would not be viable is likely to hold in more general models. The important—and robust—insight from Baxter's analysis, although it is not framed precisely this way, is that the interchange fee helps internalize an externality between the two customer groups and in so doing has the potential of making both customer groups better off.

## *Imperfect Competition in Issuing and Acquiring*

The assumption of perfect competition between homogeneous issuers and acquirers is not realistic and leads to most interesting questions having indeterminate answers—since there are no profits to be had, there is no motivation for doing anything. The next logical step was taken by Richard Schmalensee.[**********] He made the standard assumption that, except for the payment system under study, the economy was perfectly competitive.[††††††††††] This implies that the demand system facing the system, which Schmalensee took as given and did not derive from first principles, could be used for standard welfare analysis. Schmalensee allowed for imperfect competition among issuers and/or among acquirers and made a particular assumption about the functional form of the demand system and thus about the structure of indirect network effects. As in the Baxter analysis, the system itself was assumed to operate like the Visa and MasterCard systems operate in fact, on a break-even basis, and fixed costs and access prices were again

---

§§§§§§§§§§ This example is from Rochet 2003, where this model is instructively explored in more detail.

********** Schmalensee 2002.

†††††††††† Much of the subsequent literature has persuasively called that standard assumption into question in this context, as we discuss in the next section. Nonetheless, it is instructive to see its implications before considering how it might be altered in the direction of greater realism.

assumed away. Thus, as in the numerical example above, the interchange fee simply shifts costs between issuers and acquirers, raising costs on one side of the market by exactly as much as it lowers costs on the other side.

In this model the level of the interchange fee can affect the profits earned by issuing and acquiring banks whenever competition among them is imperfect, making it possible to compare profit-maximizing and welfare-maximizing fees. Under some special assumptions the comparison is simple: when there is a single issuer and a single acquirer and demand curves are linear, for instance, Schmalensee shows that the profit-maximizing interchange fee also maximizes system output and economic welfare. In this case regulation could only reduce overall performance. In general, however, these three fees may be different even under the particular demand structure Schmalensee assumed, and even when the further assumption of linear demand is imposed. Thus, even under strong assumptions about demands and costs, the relations among these quantities is complex and depends on demand parameters, cost conditions, and, an element not present in analysis of unitary two-sided platforms, the nature of competition among issuers and among acquirers.

In the Schmalensee model the interchange fee is not an ordinary market price: it is a balancing device for shifting costs between issuers and acquirers and thus shifting charges between consumers and merchants.[‡‡‡‡‡‡‡‡‡‡] Fixing the interchange fee is quite unlike fixing a price in a typical one-sided market. The first-order effect of ordinary price-fixing is to harm consumers by restricting output. The first-order effect of collective determination of interchange fees in this model is generally to enhance the value of the system by balancing participation of the two customer groups, thus internalizing indirect externalities. This is illustrated most clearly by the existence of a special case, noted above, in which collective determination of the interchange fee in order to maximize profit also maximizes output and economic welfare. And even in Schmalensee's simple model, the socially optimal interchange fee depends on costs, demand conditions, competition among issuers and among acquirers, and externalities between merchants and consumers. Thus as a practical matter, there is no rule for regulatory determination of the interchange fee that could be relied on to improve overall system performance and thus enhance economic efficiency.

[‡‡‡‡‡‡‡‡‡‡] In addition, the interchange fee can be used to shift profits between issuers and acquirers, though when it is used for that purpose total system profit is necessarily reduced.

## 4. Second-Best Interchange Fee Analysis

Except for Schmalensee, most writers on interchange assume some market distortion in addition to imperfect competition among issuers and/or acquirers. In these models social welfare analysis cannot generally be based on the demand system facing issuers and acquirers, and that demand system is derived from more fundamental assumptions rather than assumed as in the Schmalensee paper. These authors then generally examine how the additional distortion they consider affects the relation between the profit-maximizing and welfare-maximizing interchange fees. Almost all assume away fixed costs and access pricing and assume that all revenues flow from variable prices. When used to consider regulatory policy, these models become exercises in the economics of the second best.[§§§§§§§§§§] That is, they consider policy in the presence of multiple, interacting departures from the competitive ideal. In general, such exercises rarely yield tractable rules that can be used to design practical policies. The interchange literature, which points to many additional distortions but typically considers them one at a time, provides no exceptions.[**********]

### *Imperfect Competition Among Merchants*

In a very influential paper that was the first analysis of interchange to derive system demand functions from first principles, Jean-Charles Rochet and Jean Tirole allowed for imperfect competition among merchants.[††††††††††] They assumed perfect competition among acquirers and, for simplicity, identical merchants. They also assumed that some fraction of consumers was more likely to patronize merchants who accepted cards, so that merchants had a strategic incentive to accept cards in order to avoid losing the economic profits they would earn from selling to those consumers. Thus, in contrast to Baxter's model, because of this strategic,

---

§§§§§§§§§§ Lipsey and Lancaster 1956.

********** For surveys of some of the literature discussed here, see Chakravorti 2003, Hunt 2003, Rochet 2003, Rochet and Tirole 2003a, and Schmalensee 2003.

†††††††††† In an influential review of the literature undertaken for the Reserve Bank of Australia, Katz 2001 stressed the effects of such competition. See also Wright 2003c and Hayashi 2004.

rent-seeking incentive, merchants would find it optimal to accept cards even if transactions using them were somewhat more expensive than transactions using cash.

Since all profits in this model system are earned by issuers, the profit-maximizing interchange fee is the highest fee consistent with all merchants accepting the card. This may be equal to the socially optimal fee, but it may be higher because competition forces merchants to internalize part of the benefits of cardholders. In the former case card usage is welfare-maximizing, but in the latter case it is excessive, since an interchange fee that is too high drives consumer variable fees too low and thus stimulates excessive card usage by consumers.[‡‡‡‡‡‡‡‡‡‡‡] No cost-based or other simple rule for regulating the interchange fee reliably solves this problem, however, even in this bare-bones model.

The result that a profit-maximizing system never sets an interchange fee below the welfare-maximizing level in the Rochet-Tirole model, which seems to underpin some current regulatory initiatives, depends critically on the very strong assumption that merchants are identical. Julian Wright shows that if this assumption is relaxed and even if consumers know which merchants accept cards (so merchants' rent-seeking incentive to take cards is maximized), the profit-maximizing interchange fee may be above or below the welfare-maximizing level, and there may thus be too many card transactions or too few.[§§§§§§§§§§§§] The relation between profit-maximizing and welfare-maximizing fee levels in this more general model is complex and depends on details of demand and competitive conditions as well as costs.

**Reducing the Transactions Costs of Borrowing**

Credit cards bundle the provision of credit with transaction processing. In so doing they reduce the transactions costs of borrowing on the part of cardholders, so that cardholders can buy on credit more easily. Through the development of sophisticated risk scoring methods, credit cards may also have relaxed the overall liquidity constraints that many consumers face. In addition, charge cards provide short-term liquidity, albeit only for a couple of weeks on average,

[‡‡‡‡‡‡‡‡‡‡‡] It is worth noting that De Grauwe and Rinaldi 2002 have developed a model involving rent-seeking by merchants in which the level of card usage is below the welfare-maximizing level. They treat merchants' decisions to accept or not accept cards as exogenous, however.

[§§§§§§§§§§§§] Wright 2004a.

and debit and charge cards reduce the transaction costs of obtaining funds from alternative sources. If a merchant's decision to accept cards induces consumers sometimes to spend more than they otherwise would because of these reductions in consumers' transactions costs, merchants have another incentive to accept payment cards, over and above their own transaction cost savings and the rent-seeking incentive discussed above.********** There is also at least anecdotal evidence that these are benefits that merchants receive at lower cost from general purpose payment cards than they could have provided themselves through store card programs. Many smaller retailers dropped store cards as credit cards became more widely held, and few retailers steer customers to their own store cards anymore.

It is hard to evaluate the welfare implications of providing increased liquidity. Relaxing liquidity constraints and reducing transaction costs clearly benefit consumers. The more merchants that accept cards, the larger these benefits. A rigorous analysis would need also to consider the sources of both liquidity constraints and transaction costs, however, and we are not aware of any empirical or theoretical analysis that even attempts to do this.

## *Competition Among Payment Systems*

In the United States, payment card systems compete against cash and checks as well as each other. The production of cash is a government activity, subsidized through the Federal budget. And cash users do not fully internalize some of the social costs of using it, such as crime against merchants. The check system in the US is run by the Federal Reserve, which essentially forced banks early in the last century to exchange checks at par—that is, to have a zero interchange fee in the checking system.†††††††††††† The price structure in this competing system is thus not fully market-determined. Similarly, outside the United States, governments have commonly influenced the evolution of Giro (a payment system in which a bank or a post office transfers money from one account to another when it receives authorization to do so) and related transactional systems.

There is no rigorous analysis of which we are aware of the effects of government-determined pricing in these competing systems on pricing in payment card systems. It is worth

********** Chakravorti and To 2003. See also Chakravorti and Emmons 2003 and Wright 2000.

noting, for instance, that a large fraction of consumers in the United States and Western Europe pay no variable fees for writing checks, primarily as a result of the decision by banks to bundle this service (and ATM/debit cards) with the general banking relationship. This might explain why credit and charge card systems have generally not imposed transaction-specific variable fees on consumers. Similarly, there has been no analysis of which we are aware that considers the impact of this sort of competition on the relationship between profit-maximizing and welfare-optimizing interchange fee levels.

The existing literature does contain a number of (generally complex) analyses of competition between payment systems or between two-sided platforms in general.[‡‡‡‡‡‡‡‡‡‡‡‡] It is fair to say that this work is at an early stage. It seems clear that the nature of consumer and merchant behavior shapes the competitive price structure, and that there is no general tendency for competition between platforms to make the price structure closer or farther from the social optimum, though competition will generally tend to lower the price level. Suppose, for instance, that consumers all have multiple cards (this is termed "multi-homing" in the literature) and are indifferent among them, and that merchants can effectively persuade consumers which card to use. In this case competing systems will have an extra incentive to compete for merchants' favor, and this will cause a tilt in the price structure against consumers that may or may not improve performance. Similarly, if consumers tend to use only one card (to "single-home") while merchants find it easy to accept all cards, competing systems will have an extra incentive to attract consumers, and the price structure will be tilted in their favor accordingly. These aspects of consumer and retailer behavior have received little empirical study, however.[§§§§§§§§§§§§§]

## *Barriers to Surcharging*

In the United States and, at least until recently, elsewhere, payment card systems seem to have generally required merchants that accept their cards to agree not to impose a surcharge on consumers who use those cards. At first blush, this no-surcharge rule (NSR) would appear to be an artificial distortion likely to reduce performance. Dennis Carlton and Alan Frankel were the

---

[††††††††††††] Chang and Evans 2000; for an alternative view, see Frankel 1998.

[‡‡‡‡‡‡‡‡‡‡‡‡] See Armstrong 2004 and 2005, Armstrong and Wright 2004, Chakravorti and Rosson 2004, Gabszewicz and Wauthy 2004, Guthrie and Wright 2003, Manenti and Somma 2003, Rochet and Tirole 2002 and 2003b,

[§§§§§§§§§§§§§] Rysman 2004 finds that most U.S. consumers carry multiple cards but tend to concentrate their purchases on only one, suggesting a weak form of single-homing. It is unclear what combination of preferences, costs, and

first to observe that, in a fully competitive system (*a la* Baxter), if there is no NSR, and if it costs merchants nothing to charge different prices to consumers depending on what payment system they use, the interchange fee will be irrelevant and card usage will be efficient.[**********] If acquirers are perfectly competitive, an increase in the interchange fee is passed along dollar for dollar to the merchant discount, and perfect competition among merchants means that the merchant discount is passed along dollar for dollar to card-using consumers (and not at all to those who pay with cash). On the other side of the market, competition among issuers means that the increase in interchange is passed to card-users, dollar for dollar, in reduced fees. Thus card-using consumers pay the full cost of the system regardless of the interchange fee, and, if they also bear the full costs of all other payment systems, they will use payment cards if and only if they are socially less costly. Accordingly, Carlton and Frankel advocate abolishing NSRs.

It turns out that the Carlton-Frankel assumptions are stronger than necessary for the interchange fee to be irrelevant in the absence of an NSR. What is required is only costless surcharging—i.e., the ability of merchants at no cost to charge different prices depending on the means of payment used.[††††††††††††††] But without perfect competition everywhere, abolishing an NSR does not generally lead to an efficient outcome. In particular, imperfect competition among issuers then tends to lead to under-provision of card services, and merchants could use surcharges as a mechanism for price discrimination. Economic welfare may be lower than at the profit-maximizing equilibrium with an NSR—even if card usage is excessive in the latter case. In the murky realm of the second-best, this sort of ambiguity is not uncommon.

Moreover, it is clear that the assumption of costless surcharging is unrealistically strong; most merchants do not discriminate among people using different means of payment even when they are not prevented from doing so. For instance, it has generally been permissible for U.S. merchants to give a discount for cash purchases and, though this was done for a time at gasoline stations, it is now extremely rare.[‡‡‡‡‡‡‡‡‡‡‡‡‡] In the Netherlands about 10 percent of merchants imposed surcharges when they were allowed.[§§§§§§§§§§§§§] In the United Kingdom, surcharges are

---

merchant behavior this finding reflects, however, or how much a typical consumer prefers her primary card over others she carries.

[**********] Carlton and Frankel 1995a and 1995b; see also Evans and Schmalensee 1995.

[††††††††††††††] See Rochet and Tirole 2002, Wright 2003a, and, especially, Gans and King 2003b. Schwartz and Vincent 2004 provide a related analysis for a unitary system.

[‡‡‡‡‡‡‡‡‡‡‡‡‡] Chakravorti 2003, p. 55.

[§§§§§§§§§§§§§] ITM Research 2000.

permitted for credit and charge cards but surcharging is uncommon.[**********] And when NSRs were abolished in Sweden, only about 5 percent of merchants imposed surcharges.[††††††††††††††] Although we are not aware of any concrete data yet from the elimination of NSRs in Australia,[‡‡‡‡‡‡‡‡‡‡‡‡‡‡] our understanding from colleagues there is that the prevalence of surcharging, at least to date, is likely in line with these other experiences. The one instance we are aware of with a somewhat higher incidence of surcharging, although still far from pervasive, was in Denmark, where earlier this year, 19 percent of merchants (primarily grocery retailers) passed a new 0.55 Kroner debit card surcharge through to consumers.[§§§§§§§§§§§§§§]

We suspect merchants are reluctant to impose surcharges for two reasons. The first is that there are transaction costs of imposing different prices based on payment methods. The second is that consumers may prefer to patronize stores that do not surcharge. No analysis of which we are aware has considered or attempted to measure these costs and preferences. There is nothing unusual here, of course: there are many things that merchants could surcharge for—because they entail specific costs that are caused by particular customers—and do not.[**********] Parking in shopping malls is an obvious example.

## *Implications*

The quantity of recent theoretical literature discussed in this section makes it clear that economists find interchange fees fascinating. Almost all of these papers find that profit-maximizing interchange fees are unlikely to be socially optimal, but none yields workable rules for welfare-improving regulatory intervention. (In particular, none points to the optimality of any rule that is purely cost-based.) Moreover, these models are highly stylized. A variety of market imperfections are considered one at a time; not only are some visible imperfections not considered at all (e.g., government-determined zero interchange in U.S. checking), no analyst has even attempted to consider them all together. Most papers assume consumers are faced with

[**********] Office of Fair Trading 2003b, pp. 114-115.
[††††††††††††††] IMA Market Development AB 2000, p. 18; but see Katz 2001, p. 44.
[‡‡‡‡‡‡‡‡‡‡‡‡‡‡] For example, the Reserve Bank of Australia notes that "some merchants" have started surcharging but that "there are no comprehensive data." Reserve Bank of Australia 2004, p. 11.
[§§§§§§§§§§§§§§] The Danish competition authority had announced that it was going to investigate whether there had been any collective action by grocery stores. Konkurrencestyrelsen 2005.
[**********] See, for example, Evans and Schmalensee 2005, p. 131.

variable (i.e., transaction-specific) charges, even though most systems do not impose such charges and rely instead on annual fees and other access charges. Finally, there has been essentially no empirical work devoted to testing any of these models or to measuring the importance of any of the effects they predict. This literature, in short, is not very useful for either rationalizing or designing a system of interchange fee regulation. Of course, for exactly the same reasons, it is not capable of proving that the interchange fees determined by the card associations are exactly or approximately socially optimal.

It is useful to compare the results for interchange fees for payment cards with some other business practices that economists have analyzed. In many ways the theoretical results surveyed here are similar to those for advertising, research and development, product design, product variety, location decisions, firm entry in the presence of fixed costs, bundling, and price discrimination.[††††††††††††††††] In all these cases economic models show that the profit-maximizing result under imperfect competition may deviate from the social welfare-maximizing result. However, in most cases the bias can go in either direction, and in all cases determining the socially optimal result depends on complex factors that cannot be measured in practice. Based on our review of the theoretical literature on interchange fees to date there does not seem to be any basis for concluding that the potential distortions caused by collective determination of interchange fees are any more—or any less—significant than the potential distortions caused by these other deviations from the ideal model of perfect competition.

## 5. Government Determination of Interchange Fees

Traditionally, government control of prices and conditions of service—either via government ownership or economic regulation—has been most prevalent in network public utility sectors, such as water, electricity, telephone, and gas. Firms in these sectors were traditionally local monopolists. The stated purpose of regulation or government ownership in these sectors was to protect consumers from prices that would otherwise reflect the exercise of monopoly power and thus, as every student of basic economics should be able to explain, would be too high. As all of those students should also know, efficient prices in these sectors are based on marginal costs, with socially optimal markups above marginal cost depending on demand

[††††††††††††††††] See, generally, Tirole 1988.

conditions. However, the global movement toward privatization, deregulation, and incentive regulation reflected an emerging consensus that even in these near-textbook cases, economic welfare was in practice not reliably improved by government ownership or price regulation.‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ Politics inevitably intrudes into government price-setting, for instance, and limitations on profit rates tend to lead to waste and inefficiency.

Experience with government control of prices has taught analysts that a persuasive economic case for price regulation requires a positive answer to two questions:

1. Is the performance of the market or markets being considered substantially sub-optimal?

2. Is there a practical regulatory policy that is reasonably certain to improve market performance substantially?

Since regulation is a blunt policy instrument in practice, unless there is a substantial market failure there is scant chance that regulation will reliably improve matters. And unless there is a known, practical regulatory rule that, if followed, is reasonably certain to improve performance, it is likely that regulation will be on balance harmful. At least in theory, there is no economic point to interfering with even imperfect markets unless those imperfections are serious and capable of correction by known methods.

## *The First Question: Is There a Significant Market Failure?*

As we have discussed, the growing body of theoretical writing on interchange fees establishes that privately optimal fees are unlikely to be socially optimal, but it does not indicate whether they will be systematically too high or too low. This literature does not and, in the absence of evidence, cannot indicate whether non-optimal interchange fees have a significant or trivial effect on overall performance. The related literature on pricing structures for two-sided

‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ See, for instance, Armstrong et al. 1994, Megginson and Netter 2001, and Littlechild 2003. In the United States, price cap regulation has been used as an alternative to rate of return regulation by the FCC and some states in regulating telephone companies starting in the 1980s. Viscusi et al. 2000, Chapter 12. This approach has been very widely used outside the United States. Winston and Morrison 1986 document some of the adverse effects of airline regulation in the US.

markets reaches a similar result.§§§§§§§§§§§§§§§§ There is no reason to presume that even competing two-sided platforms will settle on a price structure that is socially optimal. And, for this reason, there is no reason to presume that unitary systems have pricing structures near the social optimum.

Empirical evidence, not theory, therefore must play the leading role in assessing whether interchange fees—and the resulting prices to cardholders and merchants—lead to significant underprovision or overprovision of payment card services relative to the social optimum. In a recent discussion of the rationale for regulation of interchange in Australia, I.J. Macfarlane, Governor of the Reserve Bank of Australia, similarly stresses the results of a factual inquiry rather than the deductions of economic theorists:**********

…we saw that credit cards were growing faster than the other means of payment. This was initially somewhat surprising as credit card transactions are more expensive than most other means of payment—that is, they involve a larger payment from the users of the payments system to the providers of the payments system. … Why was this possible?

Governor Macfarlane poses *almost* the right question: Are payment system usage patterns significantly inconsistent with system costs—and, it is essential to add, benefits? There is nothing unusual about a high-cost product driving out cheaper competition if the high-cost product is much better. U.S. drivers generally prefer automatic to manual transmissions in their automobiles, for instance, even though automatic transmissions cost more and are more expensive to maintain. Drivers seem to believe the difference in benefits outweighs the difference in cost.††††††††††††††††

In our context, U.S. banks typically charge consumers the same variable price for handling a check as for handling a signature-based debit card payment: zero. Debit cards are nonetheless rapidly replacing checks because consumers find them more convenient.

An approach that is based on careful measurement of costs and benefits has the potential to distinguish what is important from what is only a theoretical possibility. Dealing with quantitative evidence rather than qualitative possibilities can also inform regulatory policies if

§§§§§§§§§§§§§§§§ It is worth noting one difference explicitly. Both profit- and welfare-maximizing interchange fees depend in general on the nature of competition among issuers and among acquirers, and this additional complexity is not present in the analysis of ordinary, unitary two-sided platform businesses.

********** Macfarlane 2005.

regulation is deemed appropriate. Since the objective of interchange-fee regulation, where it is warranted, should be to correct the effects of a distortion of price signals, it is important to get a quantitative sense of the importance of those effects. Let us, accordingly, turn to the available evidence.

Some studies have argued that payment cards are used too much, based on the observation that they are more expensive on average for merchants than cash and checks but are nevertheless increasing in use at the expense of cash and checks.[‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡] While this is intuitively appealing, it ignores both cardholders and the role of benefits in determining the social optimum.

To see this, consider a simple economy with only cash and cards and with a fixed set of transactions to be executed. For each transaction, one can in principle compute the social marginal cost of executing the transaction using cash—the sum of the marginal costs to the merchant and consumer involved, as well as the net costs to all other involved parties, including governments and commercial banks. One can similarly compute the social marginal cost of executing each transaction using cards, along with the social marginal benefit of using cards as opposed to cash—the sum of the marginal benefits to consumers, merchants, governments, and commercial banks.[§§§§§§§§§§§§§§§§§] The marginal benefit in any of these cases may be negative, of course; consumers may find cards less convenient than cash for very small transactions, for instance. In this simple case the efficient outcome is clearly for cash to be used for transactions for which the social marginal cost of using cash is less than the net social marginal cost of using cards—the social marginal cost of using cards minus the social marginal benefit of using cards rather than cash—and for cards to be used for the others. Transactions-specific marginal costs and benefits are in principle necessary to assess the importance (i.e., the net social cost) of any deviation from this ideal. Considering only costs and only one of the parties involved cannot be very informative.

To our knowledge, there are no empirical studies in the literature that consider the marginal social costs and benefits for merchants and cardholders, and thus there are no

---

[††††††††††††††††] Similarly, as Joanna Stavins pointed out in a paper on electronic check truncation and presentment, bicycles may be cheaper than cars but that does not mean that society would necessarily be better off if bicycles replaced cars. Stavins 1997, p. 28.

[‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡] See, for example, Balto 2000 and Reserve Bank of Australia 2002.

[§§§§§§§§§§§§§§§§§] For consumers, benefits would be measured by changes in consumers' surplus; for the other entities the measure would generally be changes in rents.

comparisons of actual versus optimal use of alternative payment systems that have even approximate economic validity. There is some evidence on marginal costs for merchants, some highly incomplete evidence on marginal benefits for merchants, and essentially nothing on marginal costs or benefits for cardholders or other parties.**********

Garcia Swartz, Hahn, and Layne-Farrar provide a useful overview of much of the available evidence, examine some of the issues one would need to consider in evaluating whether an economy has roughly the socially optimal use of payment cards, and present some rough calculations based on available data and some plausible assumptions about consumer benefits for a few transactions sizes and types.†††††††††††††††††† They take the merchant-based cost surveys as a point of departure but then proceed to make three types of adjustments. First, merchant-based surveys of costs usually compare payment instruments at different transaction sizes, typically the average transaction size for each payment instrument. The authors depart from this norm and compare payment instruments at a set of fixed transaction sizes, small and large, to assess the sensitivity of their results. Second, by combining plausible assumptions with available data they attempt to incorporate all other parties to the transaction—consumers, the government, and commercial banks—into the calculation of cost and benefits. Finally, they consider the benefits received by the cardholder and merchants; as a result they compare the "net" cost of payment methods. Garcia Swartz, Hahn, and Layne-Farrar present rough estimates of these magnitudes and find that the optimal payment method (considering quantifiable costs and benefits) varies by transaction size, store type, and other circumstances. For many transactions it appears that credit and debit cards do not have higher "net costs" than cash or checks. While this work is instructive, however, it falls well short of a rigorous overall assessment of payment system performance.

## *The Second Question: Will Interchange Regulation Help?*

Let us assume that one has shown that it would be socially optimal to reduce the use of payment cards in favor of other payment systems. Two further questions would then need to be

********** Compare, for instance, Humphrey et al. 2003, which argues that electronic payments systems are inherently socially more efficient, and Food Marketing Institute 1998, which argues that payment cards impose excessive costs on grocers.

considered. (1) Is the interchange fee the appropriate method for trying to achieve this improvement? (A related question is whether the source of the distortion is the collective setting of the interchange fee.) (2) If the answer to that question is affirmative then do we have a method for regulating the interchange fee that is likely to increase social welfare? Since we know that interchange fees can be too high or too low, and either case can lead to a distortion, the issue is whether regulators can estimate the optimal interchange fee precisely enough to have confidence that they will improve social welfare.

The first question is more problematic than it might seem at first. As we noted earlier, card issuers charge merchants mainly variable fees and charge cardholders both access (i.e., fixed) and variable (i.e., transaction-specific) fees. The claim has been that cards are over-used because issuers—benefiting from and competing away to some extent interchange fee revenue—impose too low (possibly negative) variable fees on cardholders. Cardholders, who do not bear the costs imposed on the merchant by their choice of payment method, it is thus argued, use their cards too much.

A regulation-mandated reduction in the interchange fee tends to reduce the variable fees faced by merchants; if competition in acquiring is sufficiently intense, this reduction is one-for-one. However, reducing the interchange fee does not necessarily raise the variable fee paid by cardholders one-for-one—or, indeed, at all. Issuers can respond to the loss of interchange fee revenue by varying either fixed or variable fees, or both. The extent to which they vary each will depend on a variety of factors including the elasticity of demand for access to cards and the elasticity of demand for transactions. In addition, there are likely to be one-time costs of various sorts (including costs of changing accounting and billing systems) caused by moving from the traditional regime of zero variable charges to a regime with positive variable charges. The one-time costs of making reward programs less generous are likely to be less significant, if they exist at all.

Suppose, for instance, that issuers increased annual fees but did not reduce variable fees at all. In that extreme case the regulation of the interchange fee would not alter consumer incentives to use cards, although it might reduce the average number of cards that people carry. (Since it is hard to function in modern economies without at least one payment card, we doubt that the fraction of households with zero cards would rise noticeably.) Preliminary data from

---

[††††††††††††††††] Garcia Swartz et al. 2004.

Australia suggests that even though interchange fees were reduced by nearly half in late 2003, the marginal price to cardholders of using credit cards has not changed much. While some reward programs were made somewhat less generous, these cover only a fraction of consumers and card transactions,‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ and we have not seen widespread evidence of surcharging by merchants or the imposition of fees by issuers that increase with card usage.§§§§§§§§§§§§§§§§§ So while available data suggest that interchange fee reductions were passed through more or less completely to reductions in merchant discounts, it does not seem that the stated objective of the RBA to make consumers face the "right" variable prices for different payment methods was realized to any appreciable extent.********** Consistent with this, the data do not reveal much, if any, impact of the reforms on the use of credit cards.††††††††††††††††††

All this is not to say that dramatic reductions in the interchange fee had no economic effects. Because retailing in Australia is relatively concentrated and thus competition is likely to be imperfect,‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ it is reasonable to expect that only a fraction of the fall in merchant

‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ Industry estimates suggest that reward cards constitute about one third to one half of all credit cards in the United States; we lack comparable data for Australia. Bayot 2003.

§§§§§§§§§§§§§§§§§ Making some reward programs less generous moved some negative variable prices toward zero. Many of these changes were in the form of caps on the total number of rewards points that can be earned each year. These limits have been set at relatively high levels, over AUS$40,000 (approximately US$31,000 at the time of this writing) a year, and likely have not been a binding constraint for many cardholders. See Reserve Bank of Australia 2004, p. 11. There were also some reductions in the marginal reward per dollar spent in some cases. There is no reason to think that this would have a big effect on card volume. In fact, if one is accumulating points in a rewards program to take a trip, and the rewards program is made less generous so that more purchases are required to earn the trip, one might use the card more rather than less intensively. There would be costs of various sorts associated with imposing positive transaction fees on non-reward cards, so it is not surprising that this does not seem to have been done.

********** The RBA stated that its reforms were in line with the principle that "consumers should face prices that take into account the relative costs of producing goods and services, as well as demand conditions." Reserve Bank of Australia 2002, p. 34.

††††††††††††††††† Credit card dollar volume grew significantly in Australia in the late 1990s. The rate of growth started declining following 1999. The rate of decline in the growth rate leveled off in 2003 and 2004, in contrast to an acceleration in the rate of decline that would be expected were the reforms to have had a substantial impact. Nonetheless, the real dollar volume of credit card purchases rose 20.0 percent between 2002 and 2004, about four times as rapidly as real household consumption. Interestingly, the number of credit card accounts grew by 11.1 percent over this period, compared to a growth of only 2.6 percent in population. See Reserve Bank of Australia 2005b.

‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ Government statistics on firm concentration are not available. We have identified what data we could find from a range of sources. Many merchant categories appear to have significant levels of concentration. For example, within their respective categories, the top department store had a 71 percent share (and the top two had an 83 percent share) in 2003; the top two supermarket and grocery stores (excluding convenience, specialty and miscellaneous food stores) had a 75 percent share in 2003; the top two mobile telephone operators had a 78 percent share in 2005; the top land-line telephone operator had a 75 percent share in 2005; and the top two airlines had an 83 percent share in 2005. In contrast, the four-firm concentration index for these categories is significantly lower in the United States based on 2002 U.S. Census data: 65 percent for department stores, 33 percent for supermarkets and grocery stores, 61 percent for mobile telephone operators, and 60 percent for land-line telephone operators. The U.S.

discounts was passed on to consumers in the form of lower prices, and the remainder went to increasing retailers' profits. When consumers use cash or credit cards without rewards they are better off as a result; when they use cards with reward schemes they are likely worse off. We have seen no evidence that these effects can be associated with distinct groups of consumers – different income quintiles, for instance.

On the other side of the market, the best evidence we have been able to obtain indicates that issuers have recovered between a third and a half of the fall in interchange revenue through increased fees to consumers.§§§§§§§§§§§§§§§§§§ Thus issuers' profits have been reduced, and consumers with credit cards of all sorts have been directly harmed, particularly if they use reward cards. In order to mitigate reductions in MasterCard and Visa interchange revenues, three of the top four Australian banks have signed agreements to issue American Express or Diners Club cards, which can provide greater, unregulated transaction-related revenues to issuing banks.********** In part as a consequence, the shares of these unitary systems have risen, though from initially low levels.††††††††††††††††††††

All in all, it seems that in the short time since interchange fee reductions were imposed retailers have been made better off, issuers have been made worse off, and some consumers have been made better off (particularly those who tend to use cash a lot) and others have been make worse off (particularly those who use credit cards with reward schemes). In the long run some of these effects are likely to be undone as consumers and issuing banks move toward American Express and Diners Club, as these schemes have substantially higher average merchant discounts than the bank card systems.‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡

---

Bureau of Transportation Statistics reports the top two firms accounting for 32 percent of enplaned passengers among major U.S. airlines in 2000. The one merchant category that we found with lower concentration was for warehouse clubs and superstores, for which the four-firm concentration index in Australia was 41 percent, compared to 92 percent in the United States. See Euromonitor 2005a, Euromonitor 2005b, Euromonitor 2005c, United States Census Bureau 2004, Bureau of Transportation Statistics 2000, Maxwell 2005, and McFarland 2005.

§§§§§§§§§§§§§§§§§§ See Chang, Evans and Garcia Swartz 2005.

********** See, e.g., Cornell 2004 and 2005.

†††††††††††††††††††† Reserve Bank of Australia 2004, p. 13.

‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ Some have argued that reducing the interchange fees of the multi-party systems will put pressure on unitary systems to follow suit, but this is not persuasive. To the extent that interchange regulation disadvantages the multi-party systems, it reduces competitive pressure on the unitary systems. (We are indebted to Michael Katz for this point.) The Reserve Bank of Australia suggests that there has been some downward pressure on merchant discounts for American Express and Diners Club in Australia, with a decline of 13 basis points over the 12 months ending June 2004 (although it does not discuss whether other factors might have been responsible for this), but expresses concern that the gap between the merchant discount for the unitary versus multi-party systems had widened by 30 basis points. See Reserve Bank of Australia 2004, pp. 12-13. Given that the merchant discounts for the unitary systems were, according to one account, about 100 basis points higher than for the multi-party systems

The apparent ineffectiveness of the fairly dramatic Australian regulatory intervention in terms of its stated goals may result from the particular competitive environment in that country and the types of cards used by consumers. However, it would appear to be generally the case that the interchange fee is a highly imprecise instrument for affecting the volume of transactions on cards and thus for correcting any perceived market distortion. That is because a there is only a loose connection between interchange fees and transactions prices to cardholders.

Let us suppose for the sake of argument that regulators went further than any have gone in fact and banned the use of consumer access fees. Interchange fee regulations would then necessarily affect the transaction price paid by consumers, as they are assumed to do in the theoretical literature. (This ban would, of course, likely impose a new set of costs and distortions on the system.) There would still be the question whether it is possible to estimate the optimal interchange fee with sufficient precision that policymakers could expect to increase rather than decrease social welfare. A robust conclusion from the theoretical literature is that an estimate of the optimal interchange fee would depend on a host of factors: estimates of the price responsiveness of cardholders and merchants, indirect network effects between cardholders and merchants, competition in issuing and acquiring and among merchants, price distortions in competing payment systems, transactions costs and liquidity constraints, and marginal costs of serving cardholders and merchants. It would also depend on how competing systems—some of which may be unitary—would respond to changes in prices to cardholders and merchants. Because of the difficulty of the task, there are no serious attempts of which we are aware to estimate the socially optimal interchange fee for any real payment system. Given currently available data and estimation methods, we believe that any such attempt could at best yield highly imprecise estimates.

This task would require far more empirical information than classic public utility regulation. Most public utilities have historically been monopolies, so the strategic interaction with competitors could be ignored, unregulated prices can be safely presumed to be too high, and reducing prices until the utility just breaks even will generally increase economic welfare—at least as long as impacts on the utility's incentives for efficiency are ignored. Although

---

(Reserve Bank of Australia 2001, Figure 2.2), even if there is a modest decrease in the merchant discounts for multi-party systems attributable to the reforms, the post-reform unitary merchant discount would still exceed the pre-reform multi-party merchant discount. And the merchant discount paid on volume shifted from the multi-party to the unitary systems would increase.

sometimes present, network effects rarely played an important role in the analysis, and the calculation of optimal (Ramsey) prices required only estimates of marginal costs and demand elasticities. However, even in the public utility context it has proved difficult to calculate precise estimates of the relevant parameters, and the determination of optimal prices has often led to considerable controversy among economists and policymakers.

A robust conclusion of the economic literature on interchange fees and two-sided markets is that cost-based interchange fees are generally not socially optimal.[§§§§§§§§§§§§§§§§§§§§] Even if one were convinced, as some regulators seem to be, that current interchange fees are too high, unlike the public utility case, there is no guarantee that lowering them toward any particular target will improve welfare. In particular, there is no basis for believing that any particular cost-based formula for determining interchange fees would even provide a first approximation to the socially optimal interchange fee. Nor is there any basis at the moment for believing that moving from the collectively determined interchange fee to a fee based on any formula that considers only costs would be likely to improve social welfare. Unlike the public utility situation, therefore, there is no basis in economic theory or fact for cost-based regulation of interchange fees such as the regime adopted in Australia or by the European Commission.[**********]

In both Australia and Europe, the regulators (or at least their economic consultants) recognized that socially optimal interchange fees also depend on demand factors or network effects but, presumably, decided not to incorporate those factors because of the difficulty of doing so, citing instead the "objectivity" and "transparency" benefits of a cost-based measure.[††††††††††††††††††††] Objectivity and transparency may have benefits, of course, but by themselves they do not, necessarily lead to greater economic efficiency.

Although cost-based interchange fee regulation could by happenstance improve the efficiency of the payment system, there are two fundamental reasons to doubt that it would regularly do so in practice. The first is that regulating the interchange fee will not necessarily have a significant effect on the variable prices paid by cardholders and therefore will not

---

§§§§§§§§§§§§§§§§§§§§ The economic literature appears to be unanimous on this point; examples include Bergman 2005, Gans and King 2003a and 2003c (who accept for the purpose of argument the assumption that externalities between merchants and consumers are unimportant in mature card systems) and Wright 2003b.

********** These regulatory schemes have the further peculiarity from an economic standpoint that the merchant discount charged by unitary systems is left unregulated while the merchant discount for cooperative systems is regulated through cost-based interchange fees. This favors the unitary systems and thus leads to a further economic distortion whose effects would need to be considered in evaluating the net social benefits from moving to cost-based interchange fees.

necessarily have a significant effect on the volume of transactions—generally the putative target of the intervention. The second is that even if regulating the interchange fee could affect the variable prices paid by both merchants and cardholders, cost-based regulation is not capable of achieving the optimal prices except by happenstance. Whether cost-based regulation in practice would increase efficiency is unknown given the current state of theoretical and empirical knowledge.

## *Competition Policy and Interchange Fee Determination*

Relatively early in the development of the payment card industry, the *NaBanco* court recognized the complexity of the role played by the interchange fee and the differences between that role and the role of an ordinary price in deciding that interchange fees should be evaluated under the rule of reason rather than be subject to the *per se* condemnation of price-fixing under U.S. law. We believe the recent theoretical literature supports this approach. There is a strong economic presumption that collective determination of ordinary prices harms consumers relative to uncoordinated, competitive pricing, but there is *no* economic presumption that collectively setting interchange fees reduces output or consumer welfare as compared to any other feasible regime.

Under U.S. law, to evaluate whether the pro-competitive benefits of collectively set interchange fees outweigh the anti-competitive costs (the test under the rule of reason) one would presumably have to compare collectively set interchange fees with the results of bilateral negotiations among acquirers and issuers.[‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡] As we have discussed, bilateral negotiation would at the very least involve high transactions costs in systems, like those in the United States, with large numbers of issuers and acquirers, and it may not be feasible in such systems. But assume it is feasible, and assume away the associated transactions costs. Then one would have to examine whether bilateral negotiations would lead to lower prices and higher output than collectively set interchange fees. Existing theory does not provide much help in

---

[†††††††††††††††††††] Reserve Bank of Australia 2002, p. 12; Katz 2001, p. 29; European Commission 2002.
[‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡] That is the remedy that was sought by the plaintiff in *NaBanco*.

predicting the outcome of such bilateral negotiations,[§§§§§§§§§§§§§§§§§§§§] and, of course, the arguments in favor of interchange regulation imply that the output of credit card systems is already too high. Thus there is no reason at all to believe that bilateral negotiations would result in an interchange fee that would be closer to the social optimum than the collectively set interchange fee.

## 6. Conclusions

Economists have only scratched the surface of the theoretical and empirical work that will be needed to understand pricing in two-sided markets in general and the determination of interchange fees in particular. Like much work in economics, many of the existing theoretical models are based on highly simplified representations of the industries in question and employ highly special assumptions concerning relevant economic relationships in order to isolate individual aspects of interest. Nevertheless, several results have emerged from the literature that seem robust enough for policymakers to rely on:

1. The socially optimal prices for customer groups in multi-sided industries depend on price elasticities of demand, indirect network effects between the customer groups, marginal costs for providing goods or services to each group, and other factors.

2. Although socially optimal prices in the payment card industry depend on the same set of factors, the socially optimal interchange fee also depends on other characteristics of these industries that affect the relationship between the interchange fee and final prices. Those factors include the use of fixed and variable fees, competitive conditions among merchants, issuers, acquirers, and the nature of competition from cash, checks, and unitary payment systems.

3. Thus the socially optimal interchange fee is not in general equal to any interchange fee based on cost considerations alone.

[§§§§§§§§§§§§§§§§§§§§] As we noted above, Small and Wright 2000 conclude bilateral negotiation would threaten system viability, but it is unclear how robust this result is.

4. One cannot presume on the basis of theory alone that the interchange fee set collectively by an association is greater than, less than, or equal to the socially optimal interchange fee.

5. One cannot presume on the basis of theory alone that movements from the collectively set interchange fee to any particular cost-based interchange fee will increase or decrease social welfare.

6. One cannot presume on the basis of theory alone that the collectively set interchange fee is greater than, less than, or equal to the interchange fee that would be set by bilateral negotiations.**********

If regulators have concluded that multi-party payment systems have exhibited significantly sub-optimal economic performance, the current state of theoretical and empirical research leaves them with three serious challenges if they try to increase payment system performance overall and thereby raise economic welfare:

1. There is no empirical research that reliably addresses whether payment cards or any other payment mechanism is used too much or too little. Such research would need to consider the social costs and benefits of alternative payment systems and consider the effect of other market distortions.

2. Although it is possible that economists will be able to estimate the quantities necessary for determining optimal interchange fees, very little empirical work has been done thus far on most of the relevant quantities.

3. It is not clear that interchange fee regulation is the appropriate intervention for correcting distortions in payment systems. The interchange fee is a blunt instrument for affecting the prices faced by consumers if issuers assess fixed fees as well as variable fees.

---

********** As we noted above, in the one example of bilateral interchange negotiation on which we have seen data, the Australian EFTPOS debit card system, the fee is negative, flowing from issuers to acquirers. And the Reserve Bank of Australia is considering raising EFTPOS interchange, shifting costs from consumers to merchants. See Macfarlane 2005.

**References**


Ackerberg, Daniel A., and Gautam Gowrisankaran. 2003. "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," mimeo, University of Arizona, October.

Ahlborn, Christian, Howard H. Chang, And David S. Evans. 2001. "The Problem Of Interchange Fee Analysis: Case Without A Cause?" *European Competition Law Review*, Vol. 22, No. 8 (August), Pp. 304-312.

American Express. 2005. *American Express Company Annual Report 2004*. February.

Armstrong, Mark. 2004. "Competition In Two-Sided Markets," Mimeo, University College London.

_____. 2005. "Two-Sided Markets: Economic Theory And Policy Implications." Mimeo, University College, London, April.

_____, Simon Cowan, And John Vickers. 1994. *Regulatory Reform—Economic Analysis And British Experience*. Cambridge, Massachusetts: MIT Press.

Armstrong, Mark, And Julian Wright. 2004. "Two-Sided Markets, Competitive Bottlenecks And Exclusive Contracts," Mimeo, University College London, November.

Balto, David A. 2000. "The Problem Of Interchange Fees: Costs Without Benefits?" *European Competition Law Review*, Vol. 21, No 4 (April), Pp. 215-224.

Baxter, William F. 1983. "Bank Exchange Of Transactional Paper: Legal And Economic Perspectives," *Journal Of Law And Economics*, Vol. 26, No. 3 (October), Pp. 541-588.

Bayot, Jennifer. 2003. "Credit Card Rewards Turn From Nouns To Verbs," *New York Times*, September 21.

Bergman, Mats A. 2005. "Two-Sided Network Effects, Bank Interchange Fees, And The Allocation Of Fixed Costs." Mimeo, Swedish Competition Authority, January.

Bolt, Wilko, and Alexander F. Tieman. 2004a. "Skewed Pricing in Two-Sided Markets: An IO Approach," Working Paper No. 13, De Nederlandsche Bank, Amsterdam, October.

_____. 2004b. "A Note on Social Welfare and Cost Recovery in Two-Sided Markets," Working Paper No. 24, De Nederlandsche Bank, Amsterdam, December.

Bureau of Transportation Statistics. 2000. "Airport Activity Statistics of Certificated Air Carriers: Summary Tables 2000," Table 1, online: http://www.bts.gov/publications/airport_activity_statistics_of_certificated_air_carriers/2000/tables/table01.html.

Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken and Egg: Competition among Intermediation Service Providers," *Rand Journal of Economics*, vol. 34, no. 2 (Summer), pp. 521-552.

Carlton, Dennis W., and Alan S. Frankel. 1995a. "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, vol. 63, no. 2 (Winter), pp. 643-668.

_____, and _____. 1995b. "The Antitrust Economics of Credit Card Networks: Reply to Evans and Schmalensee," *Antitrust Law Journal*, vol. 63, no. 3 (Spring), pp. 903-915.

Chakravorti, Sujit. 2003. "Theory of Credit Card Networks: A Survey of the Literature," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 50-68.

Chakravorti, Sujit, and William R. Emmons. 2003. "Who Pays for Credit Cards?" *Journal of Consumer Affairs*, vol. 37, no. 2 (Winter), pp. 208-230.

Chakravorti, Sujit, and Alpha Shah. 2003. "Underlying Incentives in Credit Card Networks," *Antitrust Bulletin*, vol. 48, no. 1 (Spring), pp. 53-75.

Chakravorti, Sujit, and Ted To. 2003. "A Theory of Credit Cards," mimeo, Federal Reserve Bank of Chicago, July.

Chakravorti, Sujit, and Roberto Roson. 2004. "Platform Competition in Two-Sided Markets: The Case of Payment Networks," Federal Reserve Bank of Chicago, Working Paper 2004-09, July.

Chang, Howard H., and David S. Evans. 2000. "The Competitive Effects of the Collective Setting of Interchange by Payment Card Systems," *Antitrust Bulletin*, vol. 45, no. 3 (Fall), pp. 641-677.

Chang, Howard H., David S. Evans, and Daniel Garcia Swartz. 2005. "An Assessment of the Reserve Bank of Australia's Interchange Fee Reforms," Working Paper, May.

Cornell, Andrew. 2004. "Deal Highlights Change in Card Market," *Australian Financial Review*, July 12.

Cornell, Andrew. 2005. "American Express Aims for Expansion," *Australian Financial Review*, April 18.

Cruickshank, Don. 2000. "Annex D3: Interchange Fees for Credit and Debit Cards," *UK Banking: A Report to the Chancellor of the Exchequer*, UK: HM Treasury (March), pp. 247-272.

De Grauwe, Paul, and Laura Rinaldi. 2002. "A Model of the Card Payment System and The Interchange Fee," in Clemens Esser and Michael H. Stierle, eds., *Current Issues in Competition Theory and Policy*, Berlin: VWF.

Euromonitor International. 2005a. "Department Stores in Australia," October.

Euromonitor International. 2005b. "Grocery Stores, Food Retailers and Supermarkets in Australia," October.

Euromonitor International. 2005c. "Superstores and Warehouse Clubs in Australia," October.

European Commission. 2002. "Commission Decision of 24 July 2002 Relating to a Proceeding Under Article 81 of the EC Treaty and Article 53 of the EEA Agreement, Case COMP/29.373—Visa International (2002/914/EC).

_____. 2003. "Statement on Mastercard." Midday Express Press Release, October 30.

Evans, David S. 2003a. "Some Empirical Aspects of Multi-Sided Platform Industries," *Review of Network Economics*, vol. 2, issue 3 (September), 191-209.

_____. 2003b. "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation*, vol. 20, no. 2 (Summer), pp. 325-381.

Evans, David S., Andrei Hagiu, and Richard Schmalensee. 2004. "A Survey of the Economic Role of Software Platforms in Computer-Based Industries," CESifo Working Paper No. 1314, October.

Evans, David S., and Richard Schmalensee. 1995. "Economic Aspects of Payment Card Systems and Antitrust Policy toward Joint Ventures," *Antitrust Law Journal*, vol. 63, no. 3 (Spring), pp. 861-901.

_____, and _____. 2005. *Paying with Plastic, 2nd Ed.*, Cambridge, Massachusetts: MIT Press.

Food Marketing Institute. 1998. *A Retailer's Guide to Electronic Payment System Costs*, Washington, DC: Food Marketing Institute.

Frankel, Alan S., 1998. "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, vol. 66, no. 2 (Winter), pp. 313-361.

Gabszewicz, Jean J. and Xavier Y. Wauthy. 2004. "Two-Sided Markets and Price Competition with Multi-homing," mimeo, CORE, Catholic University of Louvain, Belgium.

Gans, Joshua S., and Stephen P. King. 2003a. "Approaches to Regulating Interchange Fees in Payment Systems," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 125-145.

_____, and _____. 2003b. "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis and Policy*, vol. 3, no. 1, online: http://www.bepress.com/bejeap/topics/vol3/iss1/art1/.

_____, and _____. 2003c. "A Theoretical Analysis of Credit Card Reform in Australia," *Economic Record*, v. 79, issue 247 (December), pp. 462-472.

Garcia Swartz, Daniel D., Robert W. Hahn, and Anne Layne-Farrar. 2004. "The Economics of a Cashless Society: An Analysis of the Costs and Benefits of Payment Instruments," AEI-Brookings Joint Center for Regulatory Studies, Related Publication 04-24, September.

Guthrie, Graeme, and Julian Wright. 2003. "Competing Payment Schemes," mimeo, National University of Singapore, September.

Hagiu, Andrei. 2004a. "Two-Sided Platforms: Pricing and Social Efficiency," RIETI Discussion Paper Series 04-E-035, Research Institute for the Economy, Trade and Industry, Tokyo, December.

_____. 2004b. "Two-Sided Platforms: Pricing and Social Efficiency – Extensions," RIETI Discussion Paper Series 04-E-036, Research Institute for the Economy, Trade and Industry, Tokyo, December.

Hayashi, Fumiko. 2004. "A Puzzle of Payment Card Pricing: Why Are Merchants Still Accepting Card Payments?" mimeo, Federal Reserve Bank of Kansas City, December.

Hayashi, Fumiko, Richard Sullivan, and Stuart E. Weiner. 2003. *A Guide to the ATM and Debit Card Industry*, Federal Reserve Bank of Kansas City.

HSN Consultants. 1987. *The Nilson Report*, no. 406 (June).

_____. 1991. *The Nilson Report*, no. 510 (October).

_____. 1992. *The Nilson Report*, no. 522 (April).

_____. 2001a. *The Nilson Report*, no. 737 (April).

_____. 2001b. *The Nilson Report*, no. 738 (April).

Humphrey, David, Magnus Willesson, Ted Lindblom, and Göran Bergendahl. 2003. "What Does it Cost to Make a Payment?" *Review of Network Economics*, vol. 2, no. 2 (June), pp. 159-174.

Hunt, Robert M. 2003. "An Introduction to the Economics of Payment Card Networks," *Review of Network Economics*, vol. 2, no. 2 (June), 80-96.

IMA Market Development AB. 2000. "Study Regarding the Effects of the Abolition of the Non-discrimination Rule in Sweden: Final Report, Results and Conclusions," mimeo, Lerum, Sweden, February.

ITM Research. 2000. "The Abolition of the No-discrimination Rule." mimeo, Amsterdam, March.

Jullien, Bruno. 2004. "Two-Sided Markets and Electronic Intermediaries," CESifo Working Paper No. 1345, November.

Katz, Michael L. 2001. *Reform of Credit Card Schemes in Australia II, Commissioned Report: Network Effects, Interchange Fees, and No-Surcharge Rules in the Australian Credit and Charge Card Industry*, Reserve Bank of Australia, August.

Konkurrencestyrelsen. 2005. "Konkurrencestyrelsens Undersøgelse af Dankortgebyret," ISA/1107-8900-003/bjs/fab/lkf, Copenhagen, February.

Lipsey, R.G., and Kelvin Lancaster. 1956. "The General Theory of the Second Best," *Review of Economic Studies*, vol. 24, no. 1 (October), pp. 11-32.

Littlechild, Stephen. 2003. "Reflections on Incentive Regulation," *Review of Network Economics*, vol. 2, no. 4 (December), pp. 289-315.

McFarland, Lyndal. 2005. "Update: Australia's Virgin Blue Outlook Clouded by Fuel," *CNN Money*, May 18, online: http://money.cnn.com/services/tickerheadlines/for5/200505180450DOWJONESDJONLINE000368_FORTUNE5.htm.

Macfarlane, I.J. 2005. "Gresham's Law of Payments," Talk to AIBF Industry Forum 2005, Sydney, 23 March.

Manenti, Fabio M., and Ernesto Somma. 2003. "Plastic Clashes: Competition among Closed and Open Systems in the Credit Card Industry," mimeo, Università di Padova, October.

Maxwell, Miranda. 2005. "Tough Leader Wanted for Australia's Telstra," *Reuters*, May 10, online: http://www.reuters.com/newsArticle.jhtml?type=reutersEdge&storyID=8443245.

Megginson, William L., and Jeffrey N. Netter. 2001. "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature*, vol. 39, no. 2 (June), pp. 321-389.

National Bancard Corporation v. Visa U.S.A., Inc. 1982. "Defendant's Trial Brief," Southern District of Florida, April 12.

_____. 1984. 596 F. Supp. 1231 (S.D. Fla.).

_____. 1986a. 779 F.2d 592 (11th Cir.).

_____. 1986b. 479 U.S. 923.

Office of Fair Trading. 2003a. "MasterCard Interchange Fees: Preliminary Conclusions," London, OFT634, February.

Office of Fair Trading. 2003b. "UK Payment Systems: An OFT Market Study of Clearing Systems and Review of Plastic Card Networks," London, OFT658, May.

Reserve Bank of Australia. 2001. "Reform of Credit Card Schemes in Australia I: A Consultation Document," December.

_____. 2002. "Reform of Credit Card Schemes in Australia IV: Final Reforms and Regulation Impact Statement," August.

_____. 2004. "Payment System Board Annual Report 2004."

_____. 2005a. "Reform of the EFTPOS and Visa Debit Systems in Australia: A Consultation Document," February.

_____. 2005b. "Additional Credit Card Statistics," May, online: http://www.rba.gov.au/PaymentsSystem/AustralianPaymentsSystem/ExcelFiles/RPS.xls.

Rochet, Jean-Charles. 2003. "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 97-124.

Rochet, Jean-Charles, and Jean Tirole. 2002. "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, vol. 33, no. 4 (Winter), pp. 549-570.

_____, and _____. 2003a. "An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems," *Review of Network Economics*, vol. 2, no. 2 (June), pp. 69-79.

_____, and _____. 2003b. "Platform Competition in Two-Sided Markets," *Journal of the European Economics Association*, vol. 1, no 4 (June), pp. 990-1209.

_____, and _____. 2004a. "Defining Two-Sided Markets," mimeo, IDEI, Toulouse, France, January.

_____. and _____. 2004b. "Two-Sided Markets: An Overview." Mimeo, IDEI, Toulouse, France, March.

Rysman, Marc. 2004. "An Empirical Analysis of Payment Card Usage," mimeo, Boston University, January.

Schiff, Aaron. 2003. "Open and Closed Systems of Two-Sided Networks," *Information Economics and Policy*, vol. 15, no. 4 (December), pp. 425-442.

Schmalensee, Richard. 2002. "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, vol. 50, no. 2 (June), pp. 103-122.

_____. 2003. "Interchange Fees: A Review of the Literature," *Payment Card Economics Review*, vol. 1, no. 1 (Winter), pp. 25-44.

Schwartz, Marius, and Daniel R. Vincent. 2004. "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," mimeo, Georgetown University.

Small, John, and Julian Wright. 2000. "Decentralized Interchange Fees in Open Payment Networks: An Economic Analysis," mimeo, University of Aukland, December.

Stavins, Joanna. 1997. "A Comparison of Social Costs and Benefits of Paper Check Presentment and ECP Truncation," *New England Economic Review*, (July/August), pp. 27-44.

Swedish Bankers' Association. 2004. *Banks in Sweden—Facts about the Swedish banking market*. September.

Tirole, Jean. 1988. *The Theory of Industrial Organization*, Cambridge, Massachusetts: MIT Press.

United States Census Bureau. 2004. "2002 Economic Census: Industry Series," online: http://www.census.gov/econ/census02/guide/INDSUMM.HTM.

United States Department of Commerce: Bureau of Economic Analysis. 2003 "Imputations in the National Income and Product Accounts," September 29.

_____. 2004 "Gross Domestic Product: Implicit Price Deflator," November 30.

van Raalte, Chris, and Harry Webers. 1998. "Spatial Competition with Intermediated Matching," *Journal of Economic Behavior and Organization*, vol. 34, no. 3 (March), pp. 477-488.

Vickers, John. 2005. "Public Policy and the Invisible Price: Competition Law, Regulation, and the Interchange Fee, paper presented at a Conference *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005

Visa International. 2001. *International Country Overviews*, October.

Viscusi, W. Kip, John M. Vernon, and Joseph E. Harrington, Jr. 2000. *Economics of Regulation and Antitrust*, Cambridge, Massachusetts: MIT Press.

Weiner, Stuart E., and Julian Wright. 2005. "Interchange Fees in Various Countries: Developments and Determinants," paper presented at a Conference *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005.

Winston, Clifford, and Steven A. Morrison. 1986. *The Economic Effects of Airline Deregulation*, Washington, DC: The Brookings Institution.

Wolters, Timothy. 2000. ""Carry Your Credit in Your Pocket": The Early History of the Credit Card at Bank of America and Chase Manhattan," *Enterprise and Society*, vol. 1, no. 2 (June), pp. 315-354.

Wright, Julian. 2000. "An Economic Analysis of a Card Payment Network." Mimeo, NECG and University of Auckland, December.

_____. 2003a. "Optimal Card Payment Systems," *European Economic Review*, vol. 47, no. 4 (August), pp. 587-612.

_____. 2003b. "Pricing in Debit and Credit Card Schemes," *Economics Letters*, vol. 80, no. 3 (September), pp. 305-309.

_____. 2003c. "Why Do Firms Accept Credit Cards?" mimeo, University of Auckland.

_____. 2004a. "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, vol. 52, no. 1 (March), pp. 1-26.

_____. 2004b. "One-sided Logic in Two-sided Markets," *Review of Network Economics*, vol. 3, no. 1 (March), pp. 44-64.